



UNITE DE RECHERCHE
INRIA-ROQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
BP 105
78153 Le Chesnay Cedex
France
Tél (1) 39 63 55 11

Rapports de Recherche

N° 925

Programme 5

ALGORITHME A PENALISATION SIMPLE (Méthode quadratique séquentielle non réalisable à convergence globale)

**Jean-Claude DODU
Pierre HUARD**

Novembre 1988



★ R R - 8 9 2 5 ★

Programme 5

ALGORITHME A PENALISATION SIMPLE

(Méthode quadratique séquentielle non réalisable à convergence globale)

J.C. DODU - P. HUARD

Résumé : Cette note décrit une famille d'algorithmes d'optimisation non linéaire qui utilise la programmation quadratique successive. A chaque itération, la résolution d'un programme quadratique fournit une direction de montée d'une fonction obtenue en ajoutant une pénalisation des contraintes au critère à maximiser. Pour déterminer le successeur de l'itéré courant, on peut alors effectuer une recherche linéaire avec, comme fonction de recherche, le critère d'origine ainsi pénalisé. Cette technique de pénalisation permet d'assurer la convergence globale de l'algorithme. Le but de cette note est de replacer dans un même cadre théorique les méthodes quadratiques séquentielles de ce type, en les ramenant à un algorithme très général. Ce dernier peut être considéré comme une extension de l'algorithme classique de Zangwill.

ALGORITHME WITH MERE PENALTY

(Sequential quadratic method, non-feasible and with global convergence)

J.C. DODU - P. HUARD

Abstract. This note describes a family of algorithms for nonlinear optimization, which use sequential quadratic programming. At each iteration, solving a quadratic program yields a direction of improvement of a function obtained by adding to the objective function a penalty for violating the constraints. To obtain the next iterate, a line-search can then be made, based on the penalized objective function thus obtained. This penalty technique allows global convergence of the algorithm. The aim of this note is to give a unified theoretical framework to all these methods of the sequential quadratic programming type, by viewing them as instances of a very general algorithm. The latter can be considered as an extension of the classical Zangwill's algorithm.



S O M M A I R E

	PAGES
<u>INTRODUCTION</u>	3
 1 - <u>PROBLEME POSE</u>	 8
 2 - <u>APPROXIMATION QUADRATIQUE-HYPOTHESES</u>	 8
 3 - <u>PENALISATIONS ET RECHERCHE LINEAIRE</u>	 10
3.1 - Critère pénalisé	10
3.2 - Critère pénalisé linéarisé	13
3.3 - Recherche linéaire	15
 4 - <u>ALGORITHME GENERAL MULTIVOQUE (G)</u>	 18
 5 - <u>ALGORITHME A PENALISATION SIMPLE (PS)</u>	 18
 6 - <u>CONVERGENCE DE L'ALGORITHME (PS)</u>	 20
 6.bis - <u>UTILISATION D'UNE MATRICE H VARIABLE</u>	 21
 7 - <u>ANNEXES</u>	 22
7.1 - Continuité de $P'(x)$	22
7.2 - Continuité de $\tilde{z}(x, H)$	22
7.3 - Unicité et continuité de $\tilde{v}(x, H)$	23
7.4 - Si $(x, H) \in B \times \Gamma$, alors $\tilde{z}(x, H) = x \Leftrightarrow x$ solution (réalisable) de K-T pour (P)	24
7.5 - $\tilde{\lambda}(x, r, d)$ semi-continue supérieurement	24
7.6 - Semi-continuité inférieure de λ'' , première racine positive écrétée à 1	25
7.7 - Convergence de l'algorithme général multivoque (G)	27
7.10- Propriétés de $\Phi'(y; d)$ et de $\partial \Phi[g(x)]$	30
 8 - <u>REFERENCES</u>	 32

1 - INTRODUCTION *

Cette note étudie, sur le plan théorique, la "convergence" globale des méthodes quadratiques séquentielles (méthodes Q.S.), utilisées pour la résolution des programmes à contraintes non linéaires. Le mot convergence doit être pris ici dans le sens affaibli habituel : on dit qu'un tel algorithme converge si tout point d'accumulation de la suite des itérés qu'il engendre est une solution réalisable de Kuhn-Tucker pour le problème posé.

Rappelons que les méthodes Q.S. consistent à résoudre une suite de programmes à critère quadratique et à contraintes linéaires, ces dernières étant obtenues par linéarisation, à l'itéré x^k , des contraintes du problème posé. L'itéré x^k n'est pas forcément réalisable pour ce dernier, ni la solution optimale z^k du programme quadratique attaché à x^k . Des méthodes à convergence locale ont été proposées (cf. par exemple [Ha (1976)], [D-H (1987)]), avec $x^{k+1} = z^k$: la convergence est établie si x^0 est assez proche de la solution cherchée. Mais pour la convergence globale, avec x^0 quelconque, il n'est plus possible de prendre $x^{k+1} = z^k$. Dans certaines méthodes, dites réalisables, les itérés x^k sont réalisables pour le problème posé. Le point associé z^k ne l'étant pas, du fait de la linéarisation des contraintes, il est possible de déterminer, à partir de (x^k, z^k) , un successeur x^{k+1} réalisable meilleur que x^k pour le critère, et ceci par un processus fini (cf. [Hu (1981)]). La convergence globale est alors assurée. Mais il semble difficile, pour cette classe de méthodes, d'obtenir une vitesse de convergence superlinéaire, et la liaison automatique de ces méthodes avec des méthodes à convergence locale superlinéaire ne semble pas résolue à l'heure actuelle (cf. à ce sujet l'introduction de [D-H (1986)]).

Actuellement, la littérature propose essentiellement, pour obtenir la convergence globale, d'effectuer une recherche linéaire intermédiaire sur le segment $[x^k, z^k]$, dont le résultat fournit le successeur x^{k+1} . La fonction de recherche, qui doit avoir une dérivée directionnelle positive en x^k , dans la direction $(z^k - x^k)$, est obtenue classiquement en ajoutant au critère d'origine une pénalisation, définie à partir d'une fonction concave et d'un coefficient multiplicateur positif assez grand. Précisons que l'intérêt de ces méthodes à convergence globale est qu'elles s'adaptent généralement à une vitesse asymptotique de convergence superlinéaire, du moins dans le cas de contraintes d'égalités.

* Ce travail est le fruit d'une collaboration, dans le cadre d'un contrat extra-muros EDF - INRIA

Le but de cette note, qui se limite au problème de la convergence globale, est de donner un cadre théorique synthétique de ces méthodes Q.S., en les ramenant à un algorithme très général, où la relation de récurrence est multivoque, et qui repose sur des hypothèses assez faibles. Il s'agit d'une extension de l'algorithme de Zangwill [Z (1969)], où la condition dite de stricte monotonie est fortement affaiblie : au lieu d'exiger l'existence d'une fonction numérique h telle que $h(x^{k+1}) > h(x^k)$, on n'utilise ici qu'une condition du type $h(x^{k+1}, r^k) > h(x^k, r^k)$, où r est un paramètre scalaire modifié à chaque itération.

Le modèle dit de "pénalisation simple" présenté ici pour les méthodes Q.S. contient ainsi les méthodes classiques proposées par Pschenichny [Py (1970)], Han [Ha (1977)], Mayne et Maratos [M-M (1979)], Mayne et Polak [M-P (1982)], Bertsekas [B (1982)], Bonnans et Gabay [Bo-Ga (1984)].

Nous utilisons dans ce modèle les hypothèses classiques concernant le critère et le domaine du problème posé. La fonction de pénalisation est simplement concave, et la recherche linéaire du type Armijo, peut être remplacée par une recherche linéaire "valable" plus général (cf. [Hu (1981)]₀). Enfin, pour les contraintes du problème posé, l'algorithme traite directement et indifféremment les inégalités et les égalités.

Cette note s'adresse à des non-spécialistes. Dans cette optique, de nombreux rappels sont donnés, soit par des références, soit par des annexes.

Une autre note [D-H (1988)]₀ écrite conjointement avec celle-ci, étudie un algorithme semblable, mais où la fonction de recherche linéaire est la fonction de Lagrange, sans pénalisation (algorithme du lagrangien simple). Il ne s'agit que d'une analyse des possibilités de convergence, car cette dernière n'est pas garantie, du fait que les variables duales u^k , associées aux x^k , ne sont pas obligatoirement bornées. Les rédactions de ces deux notes sont très proches, certains passages sont identiques. Cette présentation commune est volontaire, et destinée à mieux comparer l'algorithme de pénalisation simple et celui du lagrangien simple.

Un troisième algorithme, dit du lagrangien pénalisé, combine les fonctions de recherche linéaire des deux précédents.

Sa convergence est établie, mais dans le cas de contraintes en égalités seulement. Cet algorithme du lagrangien pénalisé peut se ramener, dans le cas des contraintes en égalités, à celui de pénalisation simple, étudié ici : il suffit d'imposer aux variables auxiliaires u^k , associées aux itérés x^k , une valeur nulle, $\forall k$. Il est à noter que les hypothèses faites pour le lagrangien pénalisé ne permettent plus d'assurer, comme ici, l'existence d'un point d'accumulation pour les itérés x^k . La note [D-H (1988)_c] qui l'étudie a également une présentation très proche de celle-ci.

Remarque : Le lecteur ne s'étonnera pas des discontinuités pouvant intervenir dans la numérotation des hypothèses et des annexes. En effet, une même numérotation a été adoptée pour la présente note et les deux notes [D-H (1988)_b] et [D-H (1988)_c]. Une hypothèse (ou une annexe) peut alors n'apparaître que dans l'une ou deux des trois notes, une hypothèse (ou une annexe) ne figurant que dans les notes dans lesquelles elle intervient. Ainsi l'hypothèse (H_2) n'existe que dans [D-H (1988)_b] et [D-H (1988)_c]. Elle ne figure pas (et donc n'intervient pas) dans la présente note.

NOTATIONSEnsembles

R^n	espace euclidien à n dimensions
R_+	ensemble des nombres réels positifs ou nuls
N	ensemble des entiers 0, 1, 2,...
C^1	ensemble des fonctions à valeurs réelles continuellement différentiables
$ E $	cardinal d'un ensemble E
$\overset{\circ}{E}$	intérieur d'un ensemble E
\bar{E}	fermeture d'un ensemble E
$\mathcal{P}(E)$	ensemble des parties d'un ensemble E

Vecteurs, matrices

x_i	composante d'indice i d'un vecteur colonne x
x_I	vecteur colonne de composantes x_i , $\forall i \in I$
$x.y$	produit scalaire de deux vecteurs x et y
$[x, y]$	segment fermé d'extrémités x et y
$]x, y[$	segment ouvert d'extrémités x et y
$\ x\ _\infty$	norme l_∞ d'un vecteur x : $\ x\ _\infty = \max x_i $
$\ .\ $	norme l_2 d'un vecteur ou d'une matrice
A^i	colonne d'indice i d'une matrice A
A_i	ligne d'indice i d'une matrice A
A^I	sous-matrice d'une matrice A ayant pour colonnes A^i , $\forall i \in I$
A_I	sous-matrice d'une matrice A ayant pour lignes A_i , $\forall i \in I$
tA	transposée d'une matrice A

Fonctions, dérivées

s.c.i.	semi-continue inférieurement (pour une fonction numérique)
s.c.s.	semi-continue supérieurement (pour une fonction numérique)
$O(x)$	fonction qui tend vers zéro lorsque x tend vers une limite
C^e	fonction constante sur l'ensemble de son domaine de définition
$\nabla f(x)$	gradient en x d'une fonction f de R^n dans R (vecteur colonne à n dimensions)
$\nabla g(x)$	gradient en x d'une fonction g de R^n dans R^m (matrice de dimension $m \times n$ ayant pour lignes ${}^T \nabla g_i(x)$, $i = 1, \dots, m$)
$f'(x ; y)$	dérivée directionnelle d'une fonction f de R^n dans R , au point x et dans la direction y
f'_+	dérivée à droite d'une fonction numérique f
$\partial f(x)$	sous-différentiel en x d'une fonction convexe f

Suites, limites

$\left\{ \begin{array}{l} \{x^k \mid k \in N'\} \\ \{x^k\}_{N'} \end{array} \right\}$	suite de points x^k , $k \in N' \subset \mathbb{N}$
$\left\{ \begin{array}{l} \{x^k\}_{N'} \rightarrow x \\ x^k \xrightarrow{N'} x \end{array} \right\}$	la suite $\{x^k \mid k \in N'\}$ tend vers x

$\lambda \rightarrow 0_+$ le scalaire λ tend vers zéro par valeurs positives

1 - PROBLEME POSE

On considère le programme à contraintes non linéaires suivant :

$$(P) : \quad \begin{array}{l} \text{Maximiser } f(x) \text{ sous les conditions} \\ g_i(x) \geq 0, \quad i = 1, 2, \dots, m \end{array} \quad (1.1)$$

où $f, g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ sont de classe C^1 .

On cherche à trouver une solution réalisable vérifiant les conditions de Kuhn-Tucker (K-T solution), en résolvant une suite de problèmes approchés, à critère quadratique et à contraintes linéaires.

2 - APPROXIMATION QUADRATIQUE - HYPOTHESES

Soit $Q(x, H)$ le programme quadratique suivant, où le point x et la matrice H sont des paramètres fixés :

$$Q(x, H) : \quad \begin{array}{l} \text{Maximiser } f(x) + \nabla f(x) \cdot (z-x) + \frac{1}{2} (z-x) \cdot H(z-x) \\ \text{s.l.c.} \quad g_i(x) + \nabla g_i(x) \cdot (z-x) \geq 0, \quad i = 1, \dots, m \end{array} \quad (2.0)$$

Son domaine polyédrique est désigné par $P'(x)$. On fait les hypothèses suivantes :

$$(H_1) \quad \left| \begin{array}{l} \text{Il existe un convexe compact } B \subset \mathbb{R}^n \text{ tel que} \\ B \supset \frac{0}{P'(x)} \neq \emptyset, \quad \forall x \in B \end{array} \right.$$

Remarque : Cette hypothèse implique que P' est une fonction multivoque continue sur B (cf. démonstration annexe 7.1.).

$$(H_3) \quad \left| \begin{array}{l} \text{La matrice } H \text{ est symétrique définie négative, appartenant à un} \\ \text{compact } \Gamma. \end{array} \right.$$

Remarque : Cette hypothèse implique que $Q(x, H)$ est un programme strictement "convexe" à contraintes linéaires. La solution optimale existe, car d'après (H_1) le domaine est compact. Elle est unique, d'après (H_3) . Enfin elle vérifie les conditions de Kuhn-Tucker, nécessaires et suffisantes, les contraintes étant linéaires. Ces conditions s'écrivent ici :

$$\left| \begin{array}{l} v \geq 0 \\ {}^T v \nabla g(x) + {}^T \nabla f(x) + {}^T (z-x) H = 0 \\ v \cdot [g(x) + \nabla g(x) (z-x)] = 0 \end{array} \right. \quad \begin{array}{l} (2.1) \\ (2.2) \\ (2.3) \end{array}$$

On désignera dans ce qui suit par \tilde{z} la solution optimale et \tilde{v} le (vecteur) multiplicateur de Kuhn-Tucker associé, de $Q(x, H)$.

$$\left(H_4 \right) \left| \begin{array}{l} \text{Soit} \\ E'(\tilde{z}) = \{i \mid g_i(x) + \nabla g_i(x) \cdot (\tilde{z}-x) = 0\} \\ \text{On suppose qu'il existe une constante positive } \alpha \text{ telle que} \\ y \cdot \nabla g_E(x) \cdot {}^T \nabla g_E(x) y \geq \alpha \|y\|^2, \forall (x, H) \in B \times \Gamma, \forall y \end{array} \right.$$

Ces hypothèses permettent d'établir les résultats suivants :

Proposition (2.1)

Sous (H_1) et (H_3) , $\tilde{z}(x, H)$, solution unique de $Q(x, H)$, est bornée et continue sur $B \times \Gamma$.

Démonstration : Donnée annexe 7.2.

Proposition (2.2)

Sous (H_1) , (H_3) et (H_4) , le multiplicateur $\tilde{v}(x, H)$ à l'optimum de $Q(x, H)$ est unique, borné et continu sur $B \times \Gamma$. Sa valeur est donnée par :

$$\left| \begin{array}{l} \tilde{v}_{E'} = - \left[\nabla g_E(x) \cdot {}^T \nabla g_E(x) \right]^{-1} \nabla g_E(x) [\nabla f(x) + H(\tilde{z}-x)] \\ \tilde{v}_{\bar{E}'} = 0 \end{array} \right.$$

où E' a la définition de $E'(\tilde{z})$ donnée dans (H_4)

et $\bar{E}' = \{i \mid i \notin E', i = 1, 2, \dots, m\}$

Démonstration : Donnée annexe 7.3.

Proposition (2.3)

Sous (H_1) , (H_3) et (H_4) , on a, $\forall (x, H) \in B \times \Gamma$:

$\tilde{z}(x, H) = x \Leftrightarrow x$ solution (réalisable) de K-T pour (P).

Démonstration : Donnée annexe 7.4.

Remarque : Dans tout ce qui suit, jusqu'à la section 6.3, on supposera $H = C^{te}$ et on omettra H systématiquement dans les formules. On écrira donc $Q(x)$, $\tilde{z}(x)$ et $\tilde{v}(x)$. On posera $\tilde{d}(x) = \tilde{z}(x) - x$. Par abus d'écriture, pour alléger les formules, on écrira aussi plus simplement \tilde{z} , \tilde{v} et \tilde{d} , en l'absence d'ambiguïté.

Comme il est précisé en 6.bis, l'utilisation d'une matrice H différente à chaque résolution du programme quadratique $Q(x, H)$, et arbitrairement choisie dans Γ , ne change rien quant à la nature théorique des résultats qui suivent.

3 - PENALISATIONS ET RECHERCHE LINEAIRE

3.1 - Critère pénalisé

On considère une fonction concave $\Phi : \mathbb{R}^m \rightarrow \mathbb{R}$.

Etant donné un scalaire $r > 0$, on considère une fonction $\theta_r : \mathbb{R}^n \rightarrow \mathbb{R}$ pénalisant le critère, définie par :

$$\theta_r(x) = f(x) + r \Phi[g(x)] \quad (3.1.1)$$

La dérivée directionnelle de θ_r , pour un point x et une direction d , est donnée par :

$$\theta'_r(x; d) = f'(x; d) + r \Phi'[g(x); \nabla g(x) d] \quad (3.1.2)$$

(cf. justification annexe 7.10)

Proposition (3.1.1)

Sous les hypothèses (H_1) , (H_3) , (H_4) , on a :

$$\tilde{v} \in r \partial \Phi[g(x) + \nabla g(x) \tilde{d}] \Rightarrow \theta'_r(x; \tilde{d}) \geq -\tilde{d} \cdot H \tilde{d}$$

La direction \tilde{d} est alors une direction de pente positive pour θ_r si $\tilde{d} \neq 0$.

Démonstration :

D'après (3.1.2) et (2.2), on peut écrire :

$$\begin{aligned}\theta'_r(x ; \tilde{d}) &= \nabla f(x) \cdot \tilde{d} + r \phi' [g(x) ; \nabla g(x) \tilde{d}] \\ &= -\tilde{d} \cdot H\tilde{d} + r \phi' [g(x) ; \nabla g(x) \tilde{d}] - \tilde{v} \cdot \nabla g(x) \tilde{d}\end{aligned}$$

On sait par ailleurs (éventuellement après lecture de l'annexe 7.10) que :

$\forall y \in \mathbb{R}^m, \forall d \in \mathbb{R}^n, \exists p \in \partial \phi(y)$ tel que :

$$\phi'(y ; d) = p \cdot d \quad (3.1.3)$$

Par suite :

$\exists \tilde{p} \in r \partial \phi [g(x)]$ tel que :

$$\theta'_r(x ; \tilde{d}) = -\tilde{d} \cdot H\tilde{d} + (\tilde{p} - \tilde{v}) \cdot \nabla g(x) \tilde{d} \quad (3.1.4)$$

Enfin, on a par définition des sous-gradients :

$$\tilde{p} \in r \partial \phi [g(x)] \Rightarrow \tilde{p} \cdot \nabla g(x) \tilde{d} \geq r\phi[g(x) + \nabla g(x) \tilde{d}] - r\phi[g(x)]$$

$$\tilde{v} \in r \partial \phi [g(x) + \nabla g(x) \tilde{d}] \Rightarrow -\tilde{v} \cdot \nabla g(x) \tilde{d} \geq r\phi[g(x)] - r\phi[g(x) + \nabla g(x) \tilde{d}]$$

ce qui entraîne :

$$(\tilde{p} - \tilde{v}) \cdot \nabla g(x) \tilde{d} \geq 0 \quad (3.1.5)$$

(3.1.4) et (3.1.5) achèvent la démonstration.

Remarque : Une question se pose : x étant donné, et par suite \tilde{d} et \tilde{v} aussi, existe-t-il une valeur de r permettant de satisfaire la condition :

$$\tilde{v} \in r \partial \phi [g(x) + \nabla g(x) \tilde{d}] \quad (3.1.6)$$

Cela dépend bien évidemment du choix de ϕ . Pour l'application classique :

$$\phi(y) = \sum_{i=1}^m \phi_i(y_i) , \quad \phi_i(y_i) = \min \{0, y_i\}$$

cette condition (3.1.6) s'écrit :

$$\tilde{v}_i \in r \partial \phi_i [g_i(x) + \nabla g_i(x) \cdot (\tilde{z} - x)] , \quad \forall i = 1, \dots, m.$$

Les scalaires entre crochets sont, selon i , positifs ou nuls, puisque \tilde{z} est réalisable pour $Q(x)$. En utilisant la notation $E'(\tilde{z})$, introduite dans (H_h) et représentant les indices des contraintes actives au point \tilde{z} , on peut écrire (3.1.6) sous la forme :

$$\left| \begin{array}{l} \tilde{v}_i \in [0, r] , \quad \forall i \in E'(\tilde{z}) \\ \tilde{v}_i = 0 , \quad \forall i \notin E'(\tilde{z}) \end{array} \right. \quad (3.1.6')$$

(cf. figure 1). Compte tenu des propriétés de \tilde{v} , on voit que (3.1.6) est satisfaite en prenant $r \geq \|\tilde{v}\|_\infty$.

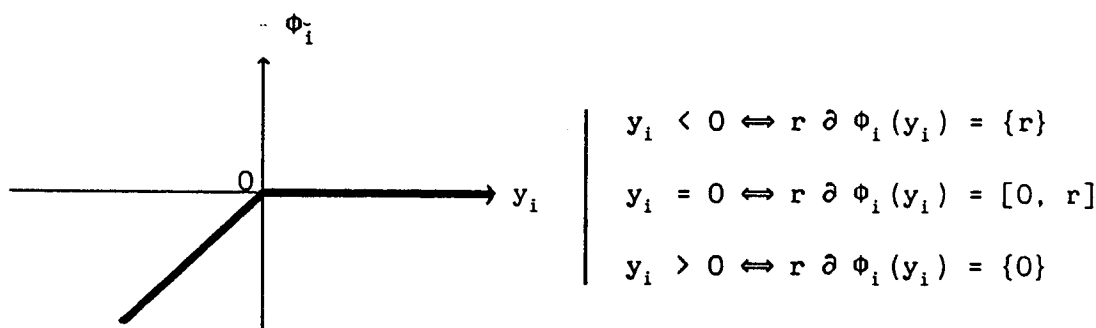


Figure 1

De même, si les contraintes de (P) constituent un système d'égalités du type $g(x) = 0$, on peut utiliser des fonctions ϕ_i définies par :

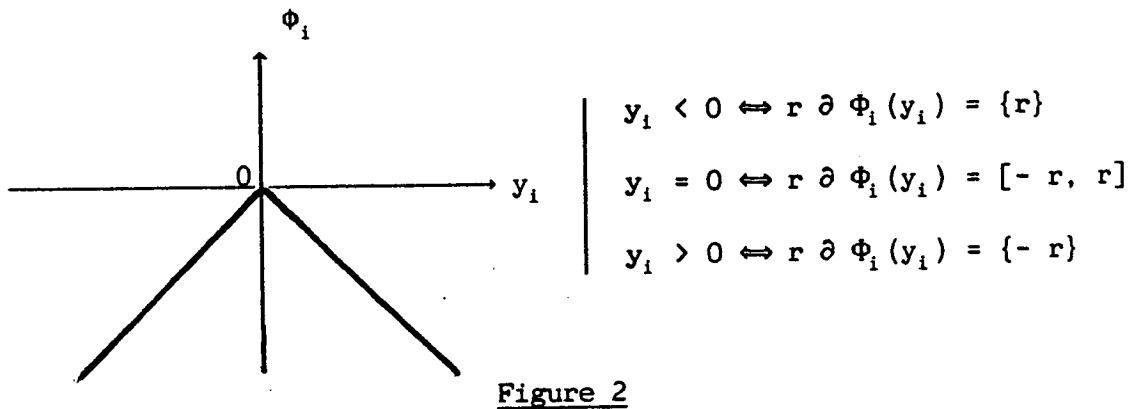
$$\phi_i(y_i) = -|y_i| , \quad \forall i = 1, \dots, m$$

En effet, on a alors, $\forall i = 1, \dots, m$ (cf. figure 2) :

\tilde{v}_i sans conditions de signe ; $\partial \phi_i(0) = [-1, +1]$;

$$g_i(x) + \nabla g_i(x) \cdot (\tilde{z} - x) = 0$$

La condition (3.1.6) est encore satisfaite avec $r \geq \|\tilde{v}\|_{\infty}$.



3.2 - Critère pénalisé linéarisé

Entre x et $\tilde{z}(x, H)$, la fonction θ_r a une dérivée directionnelle positive au point x , si r est choisi assez grand. Elle ne peut malheureusement pas servir directement pour une recherche linéaire du type Armijo, du fait qu'elle n'est pas continuellement différentiable. C'est pourquoi, au lieu de comparer cette fonction à une fonction linéaire de référence, de pente $\sigma \theta'_r(x; \tilde{d})$, avec $\sigma \in]0, 1[$, on la compare à une fonction concave $\sigma \tilde{\theta}_r(\cdot, x)$ paramétrée par x :

$$\tilde{\theta}_r(z, x) = f(x) + \nabla f(x) \cdot (z-x) + r \phi[g(x) + \nabla g(x)(z-x)]$$

On vérifie que $\tilde{\theta}_r$ est bien concave en z , x étant fixé. De plus :

$$\left\{ \begin{array}{l} \tilde{\theta}_r(x, x) = \theta_r(x) \\ \tilde{\theta}'_r(z, x; d) = f'(x; d) + r \phi'[g(x) + \nabla g(x)(z-x); \nabla g(x) d] \\ \tilde{\theta}'_r(x, x; d) = \theta'_r(x; d) \end{array} \right.$$

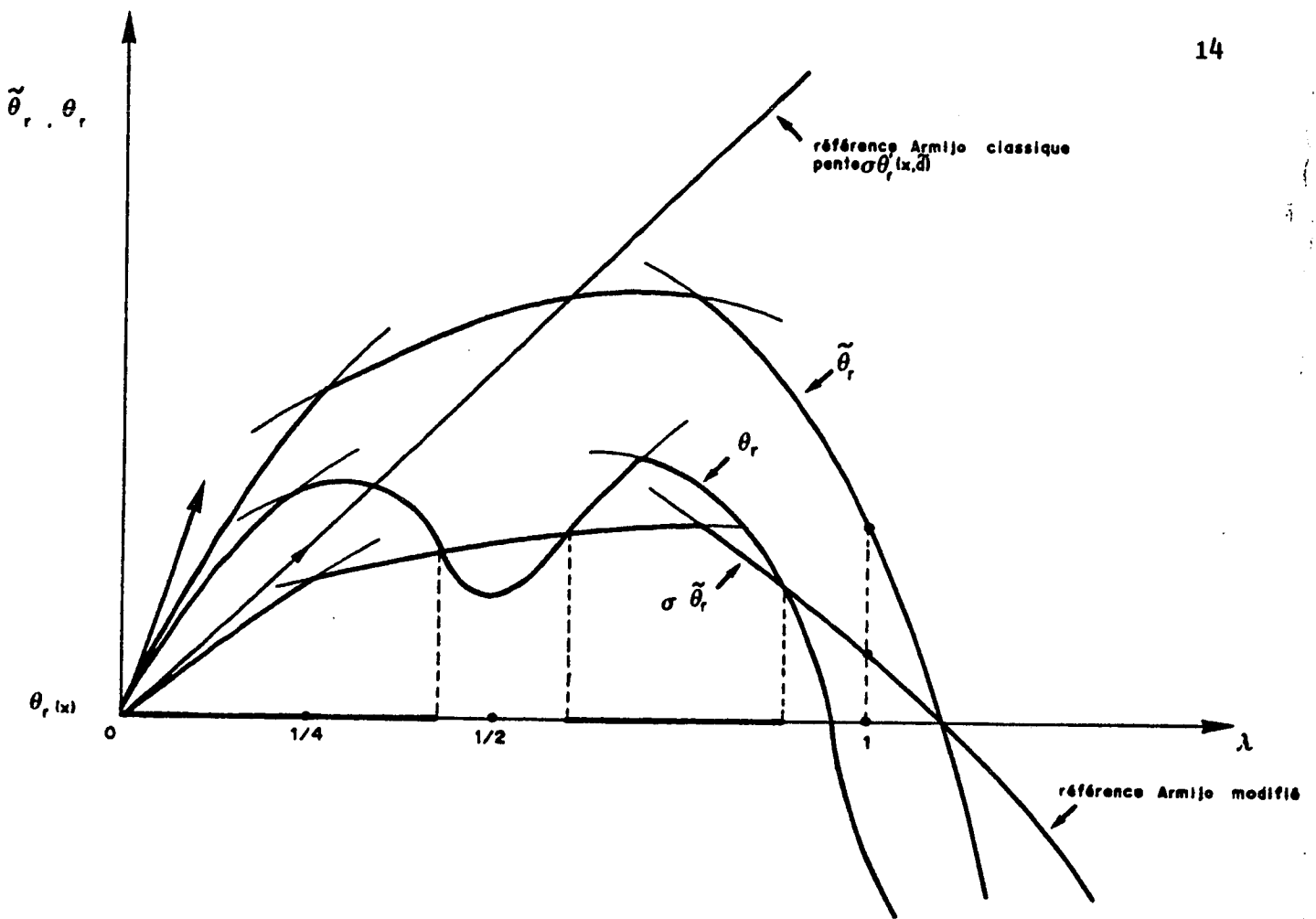


Figure 3

Proposition 3.2.1

Si $\tilde{v} \in r \partial \phi[g(x) + \nabla g(x)(\tilde{z}-x)]$ et si $\tilde{z} \neq x$, alors

$$\tilde{\theta}_r(\tilde{z}, x) - \tilde{\theta}_r(x, x) \geq -(\tilde{z}-x) \cdot H(\tilde{z}-x)$$

et par suite $\tilde{\theta}_r(\tilde{z}, x) > \tilde{\theta}_r(x, x) = \theta_r(x)$ si $\tilde{z} \neq x$

Démonstration

En posant $\tilde{d} = \tilde{z}-x$, on a :

$$\tilde{\theta}_r(\tilde{z}, x) - \tilde{\theta}_r(x, x) = \nabla f(x) \cdot \tilde{d} + r \phi[g(x) + \nabla g(x) \tilde{d}] - r \phi[g(x)]$$

$$(2.2) \Rightarrow \nabla f(x) \cdot \tilde{d} = -\tilde{d} \cdot H \tilde{d} - \tilde{v} \cdot \nabla g(x) \tilde{d}$$

$$\tilde{v} \in r \partial \phi[g(x) + \nabla g(x) \tilde{d}] \Rightarrow r \phi[g(x) + \nabla g(x) \tilde{d}] - r \phi[g(x)] \geq \tilde{v} \cdot \nabla g(x) \tilde{d}$$

$$\dots \Rightarrow \tilde{\theta}_r(\tilde{z}, x) - \tilde{\theta}_r(x, x) \geq -\tilde{d} \cdot H \tilde{d}.$$

3.3 - Recherche linéaire

Un point x , réalisable ou non pour (P), étant donné, on résout $Q(x)$. La recherche linéaire s'effectue sur le segment $[x, \tilde{z}(x)]$. Elle consiste à résoudre le problème $P(x, r, d)$ suivant, en prenant $d = \tilde{z} - x$:

Maximiser λ sous les conditions

$$P(x, r, d) : \lambda \in \Delta \quad (3.3.1)$$

$$\theta_r(x + \lambda d) - \theta_r(x) \geq \sigma [\tilde{\theta}_r(x + \lambda d, x) - \tilde{\theta}_r(x, x)] \quad (3.3.2)$$

où $\Delta = \{0\} \cup \{\lambda \mid \lambda = 2^{-i}, i \in \mathbb{N}\}$

et $\sigma \in]0, 1[$ est une constante arbitrairement choisie.

On résout $P(x, r, d)$ simplement en essayant successivement $\lambda = 1, 2^{-1}, 2^{-2}, \dots$

Il s'agit d'une modification de la recherche linéaire d'Armijo, où la fonction linéaire de référence $\lambda \theta_r'(x; d)$ est remplacée par la fonction concave en λ , $\theta_r'(x + \lambda d, x)$. Dans les deux cas, la dérivée directionnelle en x est la même pour la fonction de recherche θ_r et pour la fonction de référence. Rappelons qu'ici, cette dérivée directionnelle n'est pas continue, mais seulement s.c.i., du fait de la présence de la fonction concave ϕ .

La solution $\tilde{\lambda}(x, r, d)$ de ce problème est unique et existe toujours. Elle est positive si $\theta_r'(x; d) > 0$, car $\sigma \in]0, 1[$ et θ_r est continue. C'est le cas si $\tilde{d} \neq 0$, c'est-à-dire si x n'est pas solution réalisable de Kuhn-Tucker pour (P), d'après les propositions (3.1.1) et (2.3).

Par ailleurs, il est facile de vérifier, par une application du théorème du maximum, et du fait de la continuité de θ_r et de $\tilde{\theta}_r$, que $\tilde{\lambda}$ est une fonction semi-continue supérieurement du paramètre (x, r, d) (cf. démonstration annexe 7.5).

Proposition (3.3.1)

Si $\tilde{v} \in r \partial \phi[g(x) + \nabla g(x) \tilde{d}]$ on a :

$$\theta_r(x + \tilde{\lambda} \tilde{d}) - \theta_r(x) \geq -\sigma \tilde{\lambda} \tilde{d} \cdot H \tilde{d} \quad (3.3.3)$$

Autrement dit, le successeur de x , obtenu par la recherche linéaire, est strictement meilleur que x pour θ_r si $\tilde{z} \neq x$.

Démonstration

En posant $x' = x + \tilde{\lambda} \tilde{d}$, on a :

$$\begin{aligned} \theta_r(x') - \theta_r(x) &\geq \sigma [\tilde{\theta}_r(x', x) - \tilde{\theta}_r(x, x)] \text{ par définition de } \tilde{\lambda} \\ &\geq \sigma \tilde{\lambda} [\tilde{\theta}_r(\tilde{z}, x) - \tilde{\theta}_r(x, x)] \text{ car } \tilde{\theta}_r \text{ concave} \\ &\geq -\sigma \tilde{\lambda} \tilde{d} \cdot H\tilde{d} \text{ d'après la proposition (3.2.1)} \end{aligned}$$

Proposition (3.3.2)

Sous (H_1) , (H_3) et (H_4) , et si $\tilde{v}(x) \in r \partial \Phi[g(x) + \nabla g(x) \tilde{d}]$, on a :

$$x + \tilde{\lambda} \tilde{d} \in F_x(x, r)$$

où F_x est une fonction multivoque sup-continue, définie en tout point $(x, r) \in B \times R_+$ tel que $\tilde{z}(x) \neq x$. De plus, en de tels points :

$$x' \in F_x(x, r) \Rightarrow \theta_r(x') > \theta_r(x)$$

Démonstration

Soit $\lambda'(x, r, d)$ la première racine positive en λ , si elle existe, de l'équation :

$$\theta_r(x + \lambda d) - \theta_r(x) = \sigma [\tilde{\theta}_r(x + \lambda d, x) - \tilde{\theta}_r(x, x)] \quad (3.3.4)$$

Soit $\lambda''(x, r, d)$ cette racine "écrétée" à 1. C'est-à-dire que l'on pose $\lambda'' = \lambda'$ si λ' existe et $0 < \lambda' < 1$, et $\lambda'' = 1$ sinon. Il est clair que λ'' est toujours définie. Cette fonction λ'' est semi-continue inférieurement, comme il est montré annexe 7.6.

Par définition même de λ' , quand elle existe, on a :

$$\theta'_r(x; d) > 0 \Rightarrow \theta_r(x + \lambda d) - \theta_r(x) >$$

$$\sigma [\tilde{\theta}_r(x + \lambda d, x) - \tilde{\theta}_r(x, x)], \forall \lambda \in]0, \lambda'(x, r, d)[$$

Si $\lambda' \geq 1$, ou bien n'existe pas, la recherche linéaire donne $\tilde{\lambda} = 1$. Sinon, on vérifie aisément qu'elle donne $\tilde{\lambda} > \lambda'/2$, du fait de la contrainte $\lambda \in \Delta$.

Dans tous les cas, on a :

$$\tilde{\lambda} \geq \lambda''/2 \quad (3.3.5)$$

Posons :

$$N_1(x, r, d) = \left\{ 0 \leq \lambda \leq 1 \mid \theta_r(x + \lambda d) - \theta_r(x) \geq \sigma [\tilde{\theta}_r(x + \lambda d, x) - \tilde{\theta}_r(x, x)] \right\} \quad (3.3.6)$$

$$N_2(x, r, d) = \{\lambda \mid \lambda \geq \lambda''(x, r, d)/2\} \quad (3.3.7)$$

$$N(x, r, d) : N = N_1 \cap N_2 \quad (3.3.8)$$

En prenant :

$$F_x(x, r) = N(x, r, \tilde{d}(x))$$

la proposition est vérifiée. La sup-continuité découle de celle de N_1 (θ_r et $\tilde{\theta}_r$ sont continues), de celle de N_2 (λ'' est semi-continue inférieurement) et de la continuité de \tilde{d} , dont les valeurs sont bornées (d'après la proposition (2.1)).

L'inégalité $\theta_r(x') > \theta_r(x)$ provient du fait que :

$x' \in F_x(x, r) \Rightarrow x' = x + \lambda \tilde{d}$, avec $1 \geq \lambda \geq \lambda''/2 > 0$ et d'après la proposition (3.2.1), compte tenu de la concavité de $\tilde{\theta}_r(\cdot, x)$ et de $F_x \subset N_1$.

4 - ALGORITHME GENERAL MULTIVOQUE (G)

Proposition (4)

Soient $E \subset \mathbb{R}^n$ compact, $P \subset E$; $T \subset \mathbb{R}_+$; $\epsilon^* \in T$ fermé.

$h : E \times T \rightarrow \mathbb{R}$ continue telle que $\exists V(\epsilon^*), \exists \eta > 0$:

$$|h(x, \epsilon') - h(x, \epsilon)| \leq \eta |\epsilon' - \epsilon|, \forall \epsilon, \epsilon' \in V(\epsilon^*), \forall x \in E \quad (\alpha)$$

$G : E \rightarrow \mathcal{P}(T)$ sup-continue.

$F : E \times T \rightarrow \mathcal{P}(E)$ telle que :

$$\left| \begin{array}{l} \forall (x, \epsilon) \in E \times T, x \notin P, \epsilon \notin G(x), \text{ on a} \\ F \text{ sup-continue au point } (x, \epsilon^*) \end{array} \right. \quad (4.1.)$$

$$\left| \begin{array}{l} x' \in F(x, \epsilon) \implies h(x', \epsilon) > h(x, \epsilon) \end{array} \right. \quad (4.2.)$$

Soit $\left\{ \begin{pmatrix} x^k \\ \epsilon^k \end{pmatrix} \in E \times T \right\}_{k \in \mathbb{N}}$ une suite telle que :

$$\left| \begin{array}{l} \epsilon^k \xrightarrow{\mathbb{N}} \epsilon^*, \epsilon^{k+1} \geq \epsilon^k, \forall k \in \mathbb{N} \end{array} \right. \quad (4.3.)$$

$$\left| \begin{array}{l} x^{k+1} \in F(x^k, \epsilon^k) \text{ si } x^k \notin P \text{ et } \epsilon^k \notin G(x^k) \end{array} \right. \quad (4.4.)$$

$$\left| \begin{array}{l} x^{k+1} = x^k \text{ sinon} \end{array} \right. \quad (4.5.)$$

Dans ces conditions, si x^* est un point d'accumulation des x^k , on a :

$$\epsilon^* \notin G(x^*) \implies x^* \in P$$

La démonstration est donnée Annexe 7.7.

Remarque : Il s'agit d'une extension de l'algorithme classique de Zangwill [Z (1969)]. En particulier, l'introduction d'un paramètre ϵ dans la fonction d'itération F permet d'affaiblir la condition classique dite de "stricte monotonie", sous la forme (4.2.). On n'a donc plus obligatoirement $h(x^{k+1}, \epsilon^{k+1}) \geq h(x^k, \epsilon^k)$, $\forall k \in \mathbb{N}$.

5 - ALGORITHME A PENALISATION SIMPLE (PS)

On choisit dans cet algorithme une fonction Φ concave particulière définie par m fonctions concaves ϕ_i :

$$\Phi(y) = \sum_{i=1}^m \phi_i(y_i) \quad (y \in \mathbb{R}^m)$$

$$\text{avec } \phi_i(y_i) = \min \{0, y_i\} \quad (5.1.)$$

L'algorithme (PS) se définit alors comme suit :

Itération k

On dispose de x^k (réalisable ou non pour (P)) et de $r^{k-1} \in \mathbb{R}_+$.

$$z^k = \tilde{z}(x^k), \quad v^k = \tilde{v}(x^k), \quad d^k = \tilde{d}(x^k) = z^k - x^k$$

. Si $z^k = x^k$, STOP : x^k est solution de Kuhn-Tucker pour (P).

. Sinon, prendre :

$$r^k = \max \{r^{k-1}, \|v^k\|_\infty\}$$

$$x^{k+1} = x^k + \lambda^k d^k, \text{ avec } \lambda^k = \tilde{\lambda}(x^k, r^k, d^k)$$

On a, $\forall k \in \mathbb{N}$, $z^k \in B$ compact convexe, et $x^{k+1} \in [x^k, z^k]$. Donc $x^k \in B$ compact, $\forall k \in \mathbb{N}$, les v^k sont bornés et par suite les r^k aussi. De plus $r^k \geq r^{k-1}$. Donc $\exists r^* \in \mathbb{R}_+ : r^k \xrightarrow{\mathbb{N}} r^*$.

La mise à jour des r^k montre que l'on a, $\forall k \in \mathbb{N}$

$$v^k \in r^k \partial \Phi [g(x^k) + \nabla g(x^k) (z^k - x^k)] \quad (5.2.)$$

ce qui assure des pentes $\theta'_{r^k}(x^k; d^k)$ positives, et permet la recherche linéaire. En posant :

$$G(x) = \{r \mid \theta'_r(x; \tilde{d}(x)) \leq 0\} \quad (5.3.)$$

ou, plus explicitement :

$$G(x) = \left\{ r \mid \nabla f(x) \cdot \tilde{d}(x) + r \Phi' [g(x); \nabla g(x) \tilde{d}(x)] \leq 0 \right\}$$

on a bien évidemment pour l'algorithme (PS) :

$$r^k \notin G(x^k), \quad \forall k \in \mathbb{N} \quad (5.4.)$$

La fonction multivoque G est sup-continue, du fait de la semi-continuité inférieure de Φ' (Φ étant concave), de la continuité des autres fonctions utilisées dans la définition de G , et du sens de l'inégalité.

Par ailleurs, pour toute suite $\{(x^k, r^k)\}_{\mathbb{N}}$ engendrée par l'algorithme, et admettant un point d'accumulation (x^*, r^*) , on a, d'après la sup-continuité de $r \partial \phi [g(x) + \nabla g(x) \tilde{d}(x)]$:

$$v^* = \tilde{v}(x^*) \in r^* \partial \phi [g(x^*) + \nabla g(x^*) \tilde{d}(x^*)] \quad (5.5.)$$

6 - CONVERGENCE DE L'ALGORITHME (PS)

La suite des x^k engendrée par l'algorithme (PS) est bornée par B : son adhérence n'est pas vide. Soit x^* un tel point d'adhérence. Pour montrer que x^* est une solution de Kuhn-Tucker pour (P), nous allons vérifier que (PS) est une application de l'algorithme général (G). Il suffit de prendre dans (G) :

- . $E = B$ (B compact et $x^k \in B, \forall k \in \mathbb{N}$)
- . $P = \{x \in B \mid \tilde{z}(x) = x\}$
- . $T = R_+ ; \epsilon = r ; \epsilon^* = r^*$ (défini en 5.1.)
- . $h(x, r) = \theta_r(x)$.

On a bien $h(x, r') - h(x, r) = (r' - r) \phi [g(x)]$

$$\leq |r' - r| \max \{\phi [g(x')] \mid x' \in B\}$$

$$\forall x \in B, \forall r, r' \in T$$

- . $G(x)$ définie par (5.3.) (G est sup-continue)
- . $F = F_x$ définie Proposition (3.3.2.).

Si $x \notin P$, c'est-à-dire si $\tilde{z}(x) \neq x$, et si $r \notin G(x)$ et donc $\theta'_r(x; \tilde{d}) > 0$, on a d'après la Proposition (3.3.2.) :

- . F_x sup-continue au point (x, r) .
- . $x' \in F_x(x, r) \Rightarrow \theta'_r(x') > \theta'_r(x)$.

Le déroulement de l'algorithme conduit, si l'éventualité $x^k \in P$ ne se produit jamais, à une suite infinie telle que :

$$x^k \notin P, r^k \notin G(x^k), \forall k \in \mathbb{N}$$

La condition sur les r^k découle de la règle de mise à jour de ces derniers, cette règle impliquant successivement :

$$v^k \in r^k \partial \phi [g(x^k) + \nabla g(x^k) d^k] ; \theta'_{r^k}(x^k; d^k) > 0 ; r^k \notin G(x^k)$$

Si $r^* \notin G(x^*)$, alors d'après la proposition (4) : $x^* \in P$.

Sinon $r^* \in G(x^*)$, donc $\theta_r'(x^*; d^*) \leq 0$. Mais d'après la proposition (3.1.1.) et la propriété (5.5.) :

$$\theta_r'(x^*; d^*) \geq -d^* \cdot H d^* \geq 0$$

ce qui implique $d^* = 0$, et donc $x^* \in P$. Dans les deux cas, $x^* \in P$.

6.bis - UTILISATION D'UNE MATRICE H VARIABLE

A la fin de la section 2, une simplification a été introduite, en supposant la matrice H constante. On vérifie aisément, en reprenant les sections 3.3., 5 et 6, que l'utilisation d'une matrice variable $H^k \in \Gamma$ dans l'algorithme (PS) ne change rien sur le plan théorique. En particulier, les propriétés de continuité nécessaires pour l'utilisation du schéma général (G) sont conservées. Simplement, un triplet (x^k, H^k, r^k) est considéré au lieu de (x^k, r^k) . De même des fonctions multivoques $F_x(x, H, r)$ et $G(x, H)$ sont à prendre en considération au lieu de $F_x(x, r)$ et de $G(x)$. On a :

$$F_x(x, H, r) = N(x, r, \tilde{d}(x, H))$$

$$G(x, H) = \{r \mid \theta_r'(x; \tilde{d}(x, H)) \leq 0\}$$

Pour le programme $P(x, r, d)$, ainsi que pour sa solution $\tilde{\lambda}(x, r, d)$, la matrice H n'intervient pas comme paramètre (ne pas confondre avec $P(x, r, \tilde{d}(x, H))$).

L'application de l'algorithme général ne pose aucun problème en prenant :

$$F = F_x \times \tilde{\Gamma}$$

où $\tilde{\Gamma}$ est une fonction multivoque, dont l'image constante est le compact Γ défini section 2, hypothèse (H_3) . La suite des (x^k, H^k) engendrée par l'algorithme est bornée par $B \times \Gamma$ et admet toujours au moins un point d'accumulation (x^*, H^*) . Et quel que soit H^* , on a x^* solution de Kuhn-Tucker de (P) .

7 - ANNEXES7.1 - Continuité de $P'(x)$

La continuité de la fonction multivoque $P' : B \rightarrow \mathcal{P}(B)$ découle d'une proposition classique plus générale sur les systèmes d'inégalités paramétrées.

Proposition

Soient $X \subset \mathbb{R}^n$, $x^0 \in X$, $Z \subset \mathbb{R}^p$, Z fermé.

$\tilde{g} : X \times Z \longrightarrow \mathbb{R}^m$ continue

$\Gamma_1 : X \longrightarrow \mathcal{P}(Z)$ et $\Gamma_2 : X \longrightarrow \mathcal{P}(Z)$ définies par

$$\Gamma_1(x) = \{z \in Z \mid \tilde{g}(x, z) \geq 0\}$$

$$\Gamma_2(x) = \{z \in Z \mid \tilde{g}(x, z) > 0\}$$

On suppose :

$$(h) \quad \left| \begin{array}{l} x \in X \Rightarrow \Gamma_2(x) \neq \emptyset \text{ (et donc } \Gamma_1(x) \neq \emptyset) \\ y \in \Gamma_1(x^0) \Rightarrow y \text{ adhérent à } \Gamma_2(x^0) \end{array} \right.$$

Dans ces conditions, Γ_1 est continue au point x^0 (pour la démonstration, voir par exemple [Hu (1972)] proposition 5.8.6.).

On peut appliquer cette proposition en prenant $X = Z = B$, $\tilde{g}(x, z) = g(x) + \nabla g(x)(z-x)$ et $\Gamma_1(x) = P'(x)$. Les hypothèses (h) sont satisfaites du fait de (H_1) , et parce que $P'(x)$ est un convexe fermé d'intérieur non vide, $\forall x \in B$.

7.2 - Continuité de $\tilde{z}(x, H)$

La solution $\tilde{z}(x, H)$ du programme quadratique strictement convexe $Q(x, H)$ existe et est unique, $\forall x \in B$. Le domaine de ce programme est l'image de x par une fonction multivoque P' continue sur B (cf. 7.1.). Le critère est une fonction continue du triplet (z, x, H) . Le théorème du maximum (cf. [Hu (1972)] § 6.10.1.1) s'applique donc et garantit la sup-continuité de l'ensemble des solutions optimales. Cet ensemble étant réduit au point $\tilde{z}(x, H)$, qui demeure dans le compact B , la fonction (univoque) \tilde{z} est continue.

7.3 - Unicité et continuité de $\tilde{v}(x, H)$

A l'optimum de $Q(x, H)$, la condition (2.2.) donne $\forall(x, H) \in B \times \Gamma$:

$${}^T \nabla_{g_E} (x) \tilde{v}_E = - [\nabla f(x) + H(\tilde{z} - x)]$$

car $\tilde{v}_i = 0, \forall i \in \bar{E}'$. Par suite, par multiplication :

$$\left[\nabla_{g_E} (x) {}^T \nabla_{g_E} (x) \right] \tilde{v}_E = - \nabla_{g_E} (x) [\nabla f(x) + H(\tilde{z} - x)]$$

D'après (H_4) , la matrice multipliant \tilde{v}_E est régulière. D'où la valeur de \tilde{v} , impliquant l'unicité de \tilde{v} :

$$\left\{ \begin{array}{l} \tilde{v}_E = - \left[\nabla_{g_E} (x) {}^T \nabla_{g_E} (x) \right]^{-1} \nabla_{g_E} (x) [\nabla f(x) + H(\tilde{z} - x)] \\ \tilde{v}_{\bar{E}'} = 0 \end{array} \right. \quad (1)$$

En posant $M = \nabla_{g_E} (x) {}^T \nabla_{g_E} (x)$, où M dépend de x et de H , on vérifie aisément, avec (H_4) et la compacité de B , qu'il existe $\alpha' > \alpha$ tel que $\forall(x, H) \in B \times \Gamma$:

$$0 < \alpha \|y\|^2 \leq y.M.y \leq \alpha' \|y\|^2, \forall y$$

et que par suite :

$$0 < \frac{\alpha}{\alpha'^2} \|y\|^2 \leq y.M^{-1}.y \leq \frac{\alpha'}{\alpha^2} \|y\|^2, \forall y$$

En définitive, $\|M^{-1}\|$ est bornée, $\forall(x, H) \in B \times \Gamma$, ce qui implique que \tilde{v} est, comme \tilde{z} , borné sur $B \times \Gamma$.

Le système formé de la contrainte $g(x) + \nabla g(x)(z - x) \geq 0$ et des conditions d'optimalité (2.1.) à (2.3.) est défini par des fonctions continues de (x, H) : l'ensemble de ses solutions (\tilde{z}, \tilde{v}) est donc une fonction multivoque sup-continue.

L'unicité de (\tilde{z}, \tilde{v}) implique, avec ses bornes, que cette fonction de (x, H) est continue sur $B \times \Gamma$.

7.4 - Si $(x, H) \in B \times \Gamma$, alors $\tilde{z}(x, H) = x \Leftrightarrow x$ solution (réalisable) de K-T pour (P) (Démonstration de la Proposition (2.3.))

(i) Implication \Rightarrow

\tilde{z} vérifiant avec \tilde{v} les relations de Kuhn-Tucker (2.1.) à (2.3.), l'hypothèse $\tilde{z} = x$ implique que x vérifie les conditions de Kuhn-Tucker pour (P). De plus, $\tilde{z}(x) \in P'(x)$ par définition, et par suite $g_i(x) \geq 0$, $i = 1, \dots, m$.

(ii) Implication \Leftarrow

Soient x et v tels que :

$$v \geq 0, {}^T v \nabla g(x) + {}^T \nabla f(x) = 0, v \cdot g(x) = 0 \text{ et } g(x) \geq 0 \quad (1)$$

D'après les Propositions (2.1.) et (2.2.), il existe un couple (\tilde{z}, \tilde{v}) , solution unique en (z, v) de :

$$\left\{ \begin{array}{l} v \geq 0, {}^T v \nabla g(x) + {}^T \nabla f(x) + {}^T (z-x) H = 0 \\ v \cdot [g(x) + \nabla g(x)(z-x)] = 0, g(x) + \nabla g(x)(z-x) \geq 0 \end{array} \right\} \quad (2)$$

car les conditions de Kuhn-Tucker sont nécessaires et suffisantes pour $Q(x, H)$. Or $z = x$ est une solution de ce système (2), du fait de (1). Par suite $\tilde{z} = x$ et $\tilde{v} = v$.

Concernant les hypothèses, (i) n'utilise que (H_3) , mais pas $x \in B$. Mais (ii) utilise les hypothèses des Propositions (2.1.) et (2.2.), c'est-à-dire :

$$x \in B, (H_1), (H_3), (H_4).$$

7.5 - $\tilde{\lambda}(x, r, d)$ semi-continue supérieurement

Dans le programme $P(x, r, d)$, $\lambda \in \mathbb{R}^1$ est la variable et (x, r, d) le paramètre. Le domaine est l'intersection d'un ensemble Δ et d'une partie Π de \mathbb{R}^1 (contraintes (3.3.1.) et (3.3.2.)). Δ étant un fermé constant peut être considéré comme l'image du paramètre par une fonction multivoque (constante) sup-continue. Les fonctions θ_r et $\tilde{\theta}_r$ étant continues, Π est aussi sup-continue.

L'intersection conservant la sup-continuité, le domaine est l'image de (x, r, d) par une fonction multivoque sup-continue. Cette image demeure dans le compact Δ et le critère étant une fonction continue, l'application du théorème classique dit "du maximum" (cf. [Hu (1972)] § 6.10.1.1) est possible. Elle implique la semi-continuité supérieure de la valeur optimale $\tilde{\lambda}(x, r, d)$.

7.6 - Semi-continuité inférieure de λ'' , première racine positive écrétée à 1

Le paramètre (x, r, d) , utilisé dans la section 3, est remplacé simplement par x , ce qui ne changera rien pour les applications de la proposition qui suit.

Proposition

Soient $X \subset \mathbb{R}^n$ et $h : \mathbb{R}_+ \times X \rightarrow \mathbb{R}$ telle que

$$\left| \begin{array}{l} h \text{ continue} \\ h(0, x) = 0, \forall x \in X \end{array} \right.$$

En posant $h'_+(\lambda, x)$: dérivée à droite relative à λ , x étant fixé, on suppose que :

h'_+ existe et est s.c.i. sur $\mathbb{R}_+ \times X$

On pose $\lambda'(x)$: première racine positive en λ , si elle existe, de l'équation :

$$h(\lambda, x) = 0 \quad (x \text{ fixé})$$

et $\lambda''(x) = \lambda'(x)$ si λ' existe et si $0 < \lambda' < 1$,
 $= 1$ sinon

Soit $x^0 \in X$ tel que $h'_+(0, x^0) > 0$. Alors :

λ'' est s.c.i. au point x^0

On remarquera que λ'' est bien définie sur X .

Démonstration

Tout d'abord un lemme :

Lemme : Soit $r : R_+ \longrightarrow R$ continue et $\bar{\lambda} \in R_+$, $\bar{\lambda} > 0$, tels que :

$$\left| \begin{array}{l} r(\lambda) > 0, \forall \lambda \in]0, \bar{\lambda}[\\ r(\bar{\lambda}) = 0 \\ r'_+ \text{ existe en tout point de }]0, \bar{\lambda}[\end{array} \right.$$

Dans ces conditions, il existe une suite $\{\lambda^k \in]0, \bar{\lambda}[\}_{k \in \mathbb{N}}$ convergeant vers $\bar{\lambda}$, et telle que :

$$r'_+(\lambda^k) \leq 0, \forall k \in \mathbb{N}$$

En effet, si la conclusion du lemme n'était pas vraie, il existerait une valeur μ de $]0, \bar{\lambda}[$ telle que :

$$r'_+(\lambda) > 0, \forall \lambda \in [\mu, \bar{\lambda}[$$

La fonction continue r serait alors croissante de μ à $\bar{\lambda}$, et donc :

$$r(\bar{\lambda}) \geq r(\mu) > 0$$

en contradiction avec l'hypothèse $r(\bar{\lambda}) = 0$.

Démontrons à présent la Proposition.

Soit $\{x^k \in X\}_{k \in \mathbb{N}} \longrightarrow x^0$. Posons $\lambda^k = \lambda''(x^k)$.

Deux cas exhaustifs sont possibles :

1. $\exists N_1 \subset \mathbb{N}$, N_1 infini, tel que $0 < \lambda^k < 1$, $\forall k \in N_1$
2. $\exists \bar{k} \in \mathbb{N} : \lambda^k = 1$, $\forall k \geq \bar{k}$.

Cas 1 :

Alors $h(\lambda^k, x^k) = 0$, $\forall k \in N_1$, et λ^k est la première racine positive de l'équation. De plus $\lambda^k \in]0, 1[$, $\forall k \in N_1$. Donc $\exists N_2 \subset N_1$ tel que $\lambda^k \xrightarrow{N_2} \hat{\lambda} \in [0, 1]$. Montrons que l'on a $\hat{\lambda} > 0$.

Si $\hat{\lambda} = 0$, on peut écrire :

$$\left. \begin{array}{l} h'_+(0, x^0) > 0 \\ h'_+ \text{ s.c.i. } (\% x) \end{array} \right\} \Rightarrow \left. \begin{array}{l} \exists \bar{k} : h'_+(0, x^k) > 0, \forall k \geq \bar{k} \\ h(\lambda^k, x^k) = 0, \forall k \in N_1 \end{array} \right\} \dots$$

$$\dots \Rightarrow h(\lambda, x^k) > 0, \forall k \in N_1, k \geq \bar{k}, \forall \lambda \in]0, \lambda^k[$$

On en tire avec le lemme, et puisque h' est s.c.i. ($\% \lambda$) :

$$h'_+(\lambda^k, x^k) \leq 0, \forall k \geq \bar{k}, k \in N_1$$

ce qui implique à la limite, avec $k \in N_2$, et puisque h' est s.c.i. ($\% (\lambda, x)$) :

$$h'_+(0, x^0) \leq 0$$

en contradiction avec l'hypothèse $h'_+(0, x^0) > 0$. Donc $\hat{\lambda} > 0$. Par suite :

$$\left. \begin{array}{l} h(\lambda^k, x^k) = 0, \forall k \in N_2 \\ h \text{ continue} \end{array} \right\} \Rightarrow \left. \begin{array}{l} h(\hat{\lambda}, x^0) = 0 \\ \hat{\lambda} > 0 \end{array} \right\} \Rightarrow \lambda'(x^0) = \lambda''(x^0) \leq \hat{\lambda}$$

λ'' est bien, dans ces conditions, s.c.i. au point x^0 .

Cas 2

On a ici $\lambda''(x^k) = 1, \forall k \geq \bar{k}$, et donc

$$\lambda''(x^k) \xrightarrow{N} 1$$

Puisque $\lambda''(x^0) \in]0, 1]$ par construction, on a bien

$$1 = \lim_{k \in N} \lambda''(x^k) \geq \lambda''(x^0).$$

7.7 - Convergence de l'algorithme général multivoque (G)

Etablissons tout d'abord un lemme.

Lemme

Soit $E \subset \mathbb{R}^n$ fermé ; $T \subset \mathbb{R}_+$; $\epsilon^* \in T$.

$h : E \times T \rightarrow \mathbb{R}$ continue telle que $\exists V(\epsilon^*), \exists \eta > 0$:

$$|h(x, \epsilon) - h(x, \epsilon')| \leq \eta |\epsilon - \epsilon'|, \forall \epsilon, \epsilon' \in V(\epsilon^*), \forall x \in E \quad (\alpha)$$

On considère une suite $\{(x^k, \epsilon^k) \in E \times T\}_{k \in \mathbb{N}}$ telle que :

$$\epsilon^{k+1} \geq \epsilon^k, \forall k \in \mathbb{N}; \epsilon^k \xrightarrow{\mathbb{N}} \epsilon^* \quad (\beta)$$

$$h(x^{k+1}, \epsilon^k) \geq h(x^k, \epsilon^k), \forall k \quad (\gamma)$$

Dans ces conditions, $h(\cdot, \epsilon^*)$ est constante sur l'adhérence (si elle existe) de $\{x^k\}_{k \in \mathbb{N}}$.

Démonstration

Soient x^* et x^{**} deux points d'accumulation distincts. Il existe deux sous-suites infinies (N_1) et (N_2) , alternées, convergeant respectivement vers x^* et x^{**} .

Par ailleurs, on peut écrire (à partir d'un certain rang) :

$$\begin{aligned} h(x^{k+1}, \epsilon^{k+1}) - h(x^k, \epsilon^k) &\geq h(x^{k+1}, \epsilon^{k+1}) - h(x^{k+1}, \epsilon^k) \text{ d'après } (\gamma) \\ &\geq - |h(x^{k+1}, \epsilon^{k+1}) - h(x^{k+1}, \epsilon^k)| \\ &\geq - \eta |\epsilon^{k+1} - \epsilon^k| \quad \text{d'après } (\alpha) \\ &= - \eta (\epsilon^{k+1} - \epsilon^k) \quad \text{d'après } (\beta) \end{aligned}$$

De proche en proche et par sommation :

$$h(x^{k'}, \epsilon^{k'}) - h(x^k, \epsilon^k) \geq - \eta (\epsilon^{k'} - \epsilon^k), \forall k' \geq k.$$

Soient $k, k'' \in N_1, k' \in N_2$, tels que $k < k' < k''$.

On a :

$$h(x^{k''}, \epsilon^{k''}) \geq h(x^{k'}, \epsilon^{k'}) - \eta(\epsilon^{k''} - \epsilon^{k'}) \geq h(x^k, \epsilon^k) - \eta(\epsilon^{k''} - \epsilon^k)$$

Si $k \rightarrow \infty$, la continuité de h donne à la limite :

$$h(x^*, \epsilon^*) \geq h(x^{**}, \epsilon^*) \geq h(x^*, \epsilon^*)$$

ce qui achève la démonstration du lemme.

Revenons à la proposition (4).

Par construction, la suite des x^k est infinie et elle est bornée dans le compact E : son adhérence n'est pas vide. La proposition est vérifiée si $\exists k \in \mathbb{N}$ tel que $x^k \in P$, car alors $x^{k'} = x^k$, $\forall k' \geq k$, et $x^k \xrightarrow{\mathbb{N}} x^* \in P$.

Supposons donc $x^k \notin P$, $\forall k \in \mathbb{N}$. Soit x^* un point d'accumulation tel que $(x^*, \epsilon^*) \notin P \times G(x^*)$. Il existe $N' \subset \mathbb{N}$, N' infini, tel que :

$$(x^k, x^{k+1}, \epsilon^k) \xrightarrow{N'} (x^*, x^{**}, \epsilon^*), \text{ avec } x^{**} \in E.$$

Si $\exists k \in N' : \epsilon^{k'} \in G(x^k)$, $\forall k' \in N'$, $k' \geq k$, alors, puisque G est sup-continue, $\epsilon^* \in G(x^*)$, en contradiction avec l'hypothèse faite.

Sinon, $\exists N'' \subset N'$, N'' infini, tel que $\epsilon^k \notin G(x^k)$, $\forall k \in N''$.

On a alors :

$$(4.4.) \Rightarrow \left. \begin{array}{l} x^{k+1} \in F(x^k, \epsilon^k), \forall k \in N'' \\ x^* \notin P \\ (4.1.) \end{array} \right\} \Rightarrow \left. \begin{array}{l} x^{**} \in F(x^*, \epsilon^*) \\ x^* \notin P, \epsilon^* \notin G(x^*) \\ (4.2.) \end{array} \right\} \Rightarrow h(x^{**}, \epsilon^*) > h(x^*, \epsilon^*)$$

en contradiction avec le lemme. Donc $x^* \in P$.

7.10- Propriétés de $\Phi'(y; d)$ et de $\partial \Phi[g(x)]$ (cf. section 3.1.)Proposition

Soient $\Phi : \mathbb{R}^m \rightarrow \mathbb{R}$ concave
 $g : \mathbb{R}^n \rightarrow \mathbb{R}^m, g \in C^1$
 $h : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par $h = \Phi \circ g$.

Alors, pour tout $x \in \mathbb{R}^n, y \in \mathbb{R}^m, d \in \mathbb{R}^n, d \neq 0$, on a :

$$(i) \quad \exists p \in \partial \Phi(y) : \Phi'(y; d) = p \cdot d$$

$$(ii) \quad h'(x; d) = \Phi'[g(x); \nabla g(x) d]$$

Démonstration

(i) Propriété classique :

Φ étant concave, on sait que :

$$\Phi'(y; d) = \min_{p'} \{p' \cdot d \mid p' \in \partial \Phi(y)\}$$

$\partial \Phi(y)$ étant compact, $\forall y$, le minimum est atteint pour un certain élément de $\partial \Phi(y)$.

(ii) [F (1987)]

On a par définition :

$$h'(x; d) = \lim_{\lambda \rightarrow 0_+} \frac{\Phi[g(x + \lambda d)] - \Phi[g(x)]}{\lambda}$$

Par ailleurs, puisque $g \in C^1$:

$$g(x + \lambda d) = g(x) + \lambda \nabla g(x) d + \lambda o(\lambda)$$

En posant :

$$\delta_1(\lambda) = \frac{\Phi[g(x) + \lambda \nabla g(x) d + \lambda o(\lambda)] - \Phi[g(x) + \lambda \nabla g(x) d]}{\lambda}$$

$$\delta_2(\lambda) = \frac{\Phi[g(x) + \lambda \nabla g(x) d] - \Phi[g(x)]}{\lambda}$$

on peut écrire :

$$h'(x; d) = \lim_{\lambda \rightarrow 0_+} [\delta_1(\lambda) + \delta_2(\lambda)]$$

Φ étant concave, elle admet une dérivée directionnelle. D'où :

$$\lim_{\lambda \rightarrow 0_+} \delta_2(\lambda) = \Phi' [g(x); \nabla g(x) d] \quad (1)$$

Reste à montrer que $\lim_{\lambda \rightarrow 0_+} \delta_1(\lambda) = 0$. Posons :

$$\begin{cases} y^1(\lambda) = g(x) + \lambda \nabla g(x) d \\ y^2(\lambda) = y^1(\lambda) + \lambda 0(\lambda) \end{cases}$$

On sait, par définition même du sous-différentiel de Φ , que :

$\forall p^1 \in \partial \Phi(y^1), \forall p^2 \in \partial \Phi(y^2)$, on a

$$p^2 \cdot (y^2 - y^1) \leq \Phi(y^2) - \Phi(y^1) \leq p^1 \cdot (y^2 - y^1)$$

En appliquant ce résultat à $y^1(\lambda)$ et $y^2(\lambda)$, on obtient, $\forall \lambda > 0$:

$\forall p^1 \in \partial \Phi(y^1(\lambda)), \forall p^2 \in \partial \Phi(y^2(\lambda))$, on a

$$p^2 \cdot 0(\lambda) \leq \delta_1(\lambda) \leq p^1 \cdot 0(\lambda) \quad (2)$$

En faisant tendre λ vers zéro, les points $y^1(\lambda)$ et $y^2(\lambda)$ tendent vers $g(x)$.

Pour λ assez proche de zéro, ils appartiennent tous les deux à une boule de centre $g(x)$. L'image d'un compact par un sous-différentiel étant bornée, il existe un nombre $\eta > 0$ tel que :

$$\|p^1\| \leq \eta, \quad \|p^2\| \leq \eta$$

Par suite, (2) implique que $\delta_1(\lambda) \longrightarrow 0$ quand $\lambda \rightarrow 0$.

8 - REFERENCES

- [B (1982)] D.P. BERTSEKAS
Constrained optimization and Lagrange multiplier methods.
Academic Press, New-York
- [Bo - Ga (1984)] J.F. BONNANS et D. GABAY
Une extension de la programmation quadratique successive.
Lecture Notes in Control and Information Sciences, n° 63,
Springer-Verlag.
- [D-D (1979)] J. DELAHAYE et J. DENEL
The continuities of the point-to-set maps. Definitions and
equivalences. Dans Mathematical Programming Study n° 10,
p. 8-12, North-Holland.
- [D-H (1986)] J.C. DODU et P. HUARD
La méthode de Newton sous contraintes convexes : un
algorithme à convergence globale superlinéaire. Bulletin
de la Direction des Etudes et Recherches, Electricité de
France, Série C, n° 3, p. 31-56.
- [D-H (1987)] J.C. DODU et P. HUARD
Méthodes de quasi-Newton sous contraintes non linéaires :
algorithmes à convergence locale superlinéaire. Note
Electricité de France HR 30.0884.
- [D-H (1988)_b] J.C. DODU et P. HUARD
Algorithme du lagrangien simple. Note Electricité de
France à paraître.
- [D-H (1988)_c] J.C. DODU et P. HUARD
Algorithme du lagrangien pénalisé. Note Electricité de
France à paraître.
- [F (1987)] J.F. FAUGERAS
Conditions nécessaires d'optimalité pour une classe de
problèmes d'optimisation non différentiable. Note
Electricité de France HR 34.1027.
- [Ha (1976)] S.P. HAN
Superlinearly convergent variable metric algorithms for
general nonlinear programming problems. Mathematical
Programming, Vol 11, n° 3, p. 263-282.
- [Ha (1977)] S.P. HAN
A globally convergent algorithm for nonlinear programming.
Journal of Optimization Theory and Applications, Vol 22,
p. 297-309.

- [Hu (1972)] P. HUARD
Optimisation dans R^n . Cours polycopié du Laboratoire de Calcul de l'Université de Lille.
- [Hu (1981)_a] P. HUARD
Feasible variable metric method for nonlinearly constrained problems. Lecture Notes in Control and Information Sciences, n° 30, Springer-Verlag.
- [Hu (1981)_b] P. HUARD
Une approche synthétique des recherches linéaires dans les méthodes de pentes. Bulletin de la Direction des Etudes et Recherches, Electricité de France, Série C, n° 2, p. 89-132.
- [Hu (1982)] P. HUARD
Un algorithme général de gradient réduit. Bulletin de la Direction des Etudes et Recherches, Electricité de France, Série C, n° 2, p. 91-109.
- [M-M (1979)] D.Q. MAYNE et N. MARATOS
A first order exact penalty function algorithm for equality constrained optimization problems. Mathematical Programming, Vol 16, p. 303-324.
- [M-P (1982)] D.Q. MAYNE et E. POLAK
A superlinearly convergent algorithm for constrained optimization problems. Dans Mathematical Programming Study n° 16, p. 45-61, North-Holland.
- [Py (1970)] B.N. PSCHENICHNY
Algorithms for the general problem of mathematical programming. Kibernetika (Kiev), Vol 6, p. 120-125.
- [Z (1969)] W.I. ZANGWILL
Nonlinear programming : a unified approach. Prentice-Hall, Englewoods Cliffs, N.J.

E.D.F.

Service A.G.T.

SITE DE CLAMART

Pour tous renseignements :

Bâtiment X Bureau S/S 105

Tél. : 38.84 ou 44.98

REPROGRAPHIE

Nom, Prénom : LE ROY Bernadette COUNAN Agnès		Service : <u>ETUDES DE RESEAUX</u>		
N° de Bureau : C 312		Département : <u>METHODES D'OPTIMISATION</u>		
N° de Téléphone : 39.65 ou 34.44		Imputation : <u>R00R16</u>		
Délai :		Date :		
ORIGINAL		TIRAGE		
Nombre de pages	Format	Nombre d'exemplaires	Nature	Observations

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique

